Context-free grammar (CFG)

汇报人：扣扣

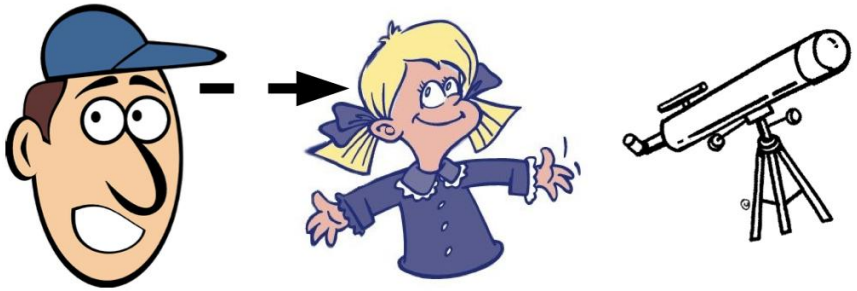
大纲

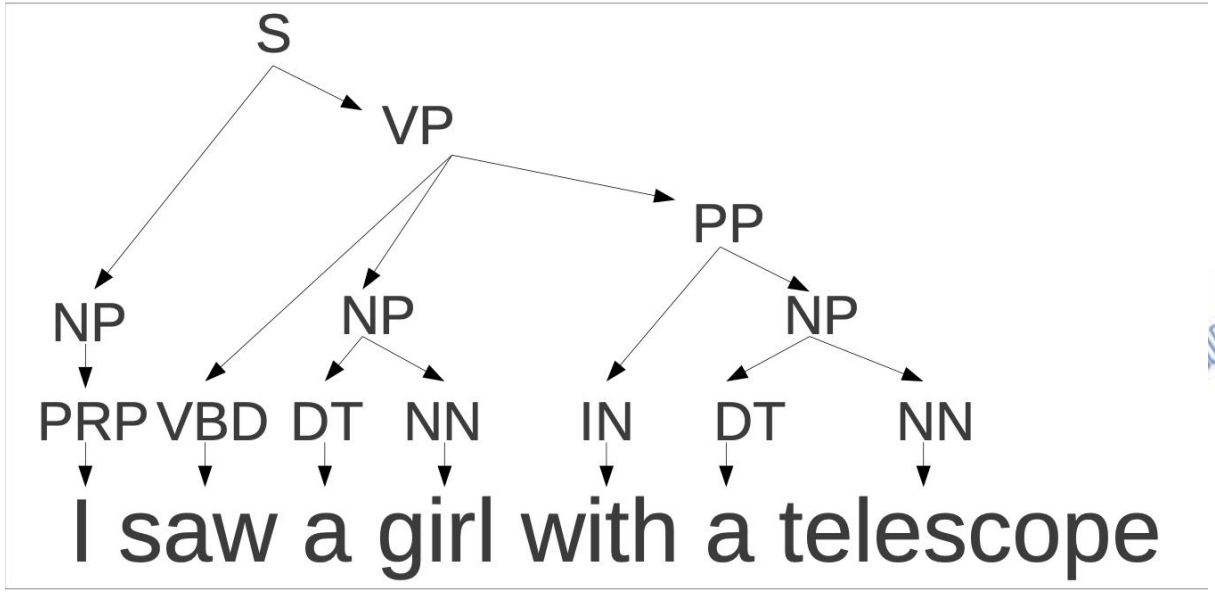
- 应用
- Context-free grammar 基本概念
- 如何理解上下文无关
- 歧义性
- 如何解决歧义性
- Chomsky Normal Form
- CKY Parsing

艾弗拉姆·诺姆·乔姆斯基（Avram Noam Chomsky，1928年12月7日—），美国哲学家，是麻省理工学院语言学的荣誉退休教授。乔姆斯基的《句法结构》被认为是20世纪理论语言学研究上最伟大的贡献。

应用

- 句子结构
- 信息抽取
- 问答系统





贪心科技 GREEDY TECHNOLOGY
京东智联云 &
nks All

Context-free grammar 基本概念

Context-free grammar：

即上下文无关文法，是一种形式文法（formal grammar）。形式文法是形式语言（formal language）的文法，由一组产生规则（production rules）组成，描述该形式语言中所有可能的字符串形式。

形式文法一般可分为四大类：无限制文法（unrestricted grammars），上下文相关文法，上下文无关文法和正则文法（Regular grammar）。

NeuHub AI Links All

Context-free grammar 基本概念

Terminal symbols： 终结符，可以理解为基础符号，词法符号，是不可替代的，天然存在，不能通过文法规则生成。

Nonterminal symbols： 非终结符，或者句法变量。

Production rules： grammar 是由终结符集、非终结符集和产生规则共同组成。产生规则定义了符号之间如何转换替代。规则的左侧是规则头，是可以被替代的符号；右侧是规则体，是具体的内容。

Context-free grammar 基本概念

例：

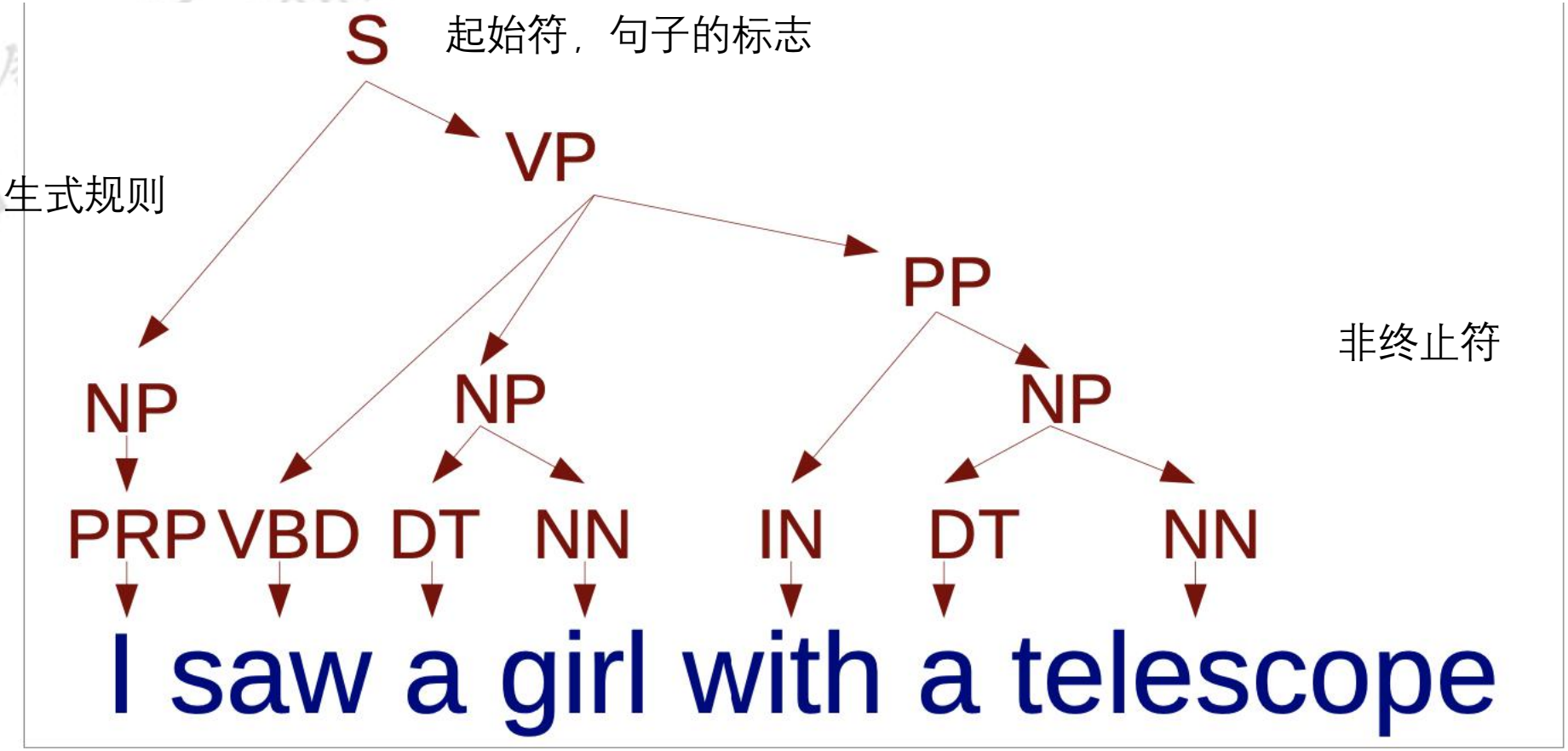
S -> cAd

A -> a|ab

- (a, b, c, d) 为终结符（用小写表示），(S, A) 为非终结符（用大写表示）。
- S -> cAd， A -> a|ab表示产生式规则。
- 可以产生"cad"，“cabd” 等符合文法的内容。

NeuHub AI Links All

Context-free grammar 基本概念



如何理解上下文无关

在应用一个产生式进行推导时，前后已经推导出的部分结果（终止符）就是上下文。

上下文无关的意思：只要文法的定义里有某个产生式，不管一个非终结符前后文是什么，就可以应用相应的产生式进行推导。

从形式上来看，就是产生式的左边都是单独一个非终结符，即形如 S-> ...，而不是非终结符左右还有别的东西，例如 aSb -> ...

NeuHub AI Links All

如何理解上下文无关

上下文无关的产生式：

S -> ABC

A -> 人 | 天

B -> 吃 | 下

C -> 雨 | 雪 | 饭 | 肉

英文字母都是非终结符（ABC 分别表示主谓宾），汉字都是终结符。

这个文法可以生成如下句子｛人吃饭，天下雨，人吃肉，天下雪，人下雪，天下饭，天吃肉，……｝

其中有一些搭配在语义上并不恰当：例如“天吃肉”

其推导过程为：S -> ABC -> 天BC -> 天吃C -> 天吃肉

京东智联云 & AI
NeuHub AI Links All

如何理解上下文无关

上下文相关的产生式：

S -> ABC

A -> 人 | 天

人B -> 人吃

天B -> 天下

下C -> 下雨 | 下雪

吃C -> 吃饭 | 吃肉

对 V 的推导过程施加了约束：虽然 V 还是能推出”吃“和”下“两个词，但是仅仅当 V 左边是”人“时，才允许它推导出”吃“；而当 V 左边是”天“时，允许它推导出”下“。这样通过上下文的约束，就保证了主谓搭配的一致性。类似地，包含 O 的产生式也约束了动宾搭配的一致性。

如何理解上下文无关

在应用一个产生式进行推导时，前后已经推导出的部分结果（终止符）就是上下文。

上下文无关的意思：只要文法的定义里有某个产生式，不管一个非终结符前后文是什么，就可以应用相应的产生式进行推导。

从形式上来看，就是产生式的左边都是单独一个非终结符，即形如 S-> ...，而不是非终结符左右还有别的东西，例如 aSb -> ...

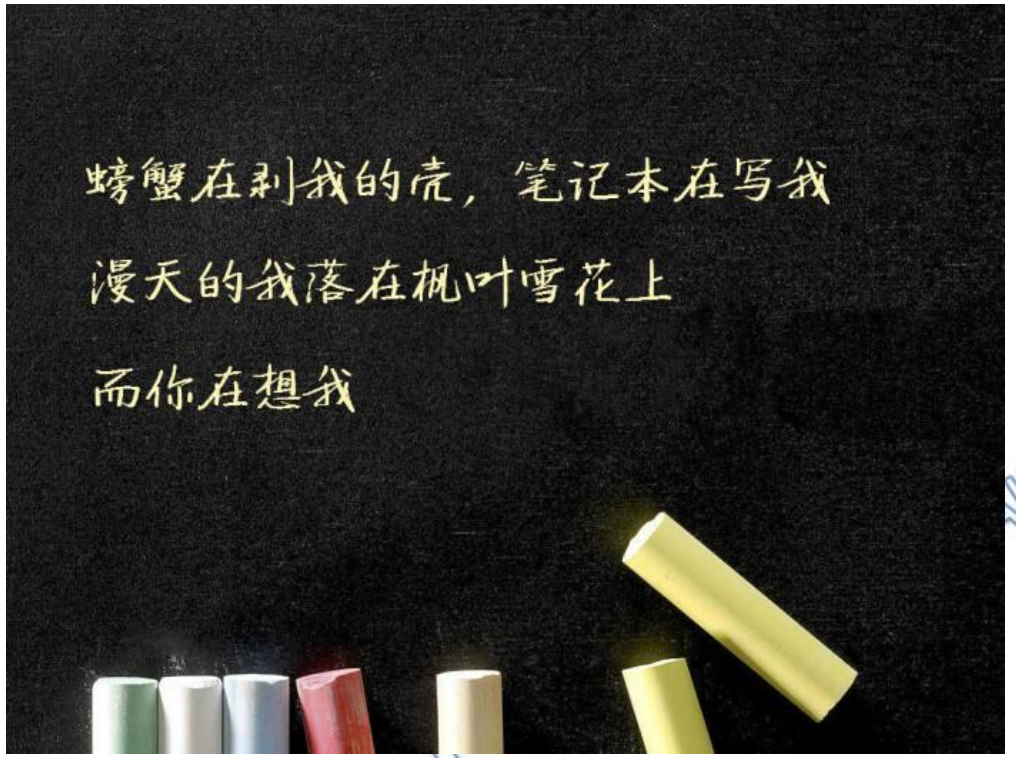
NeuHub AI Links All

如何理解上下文无关

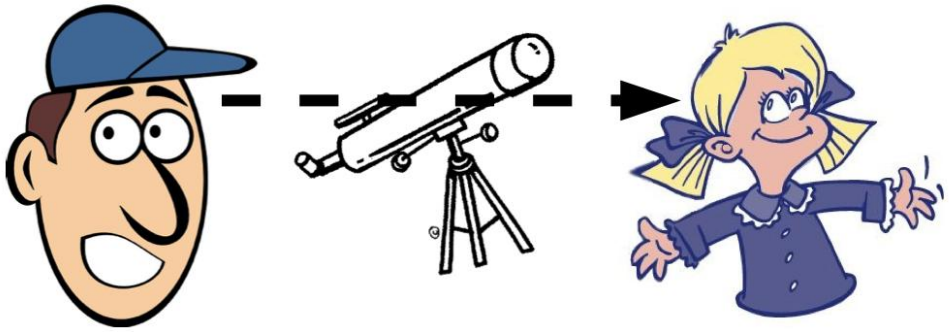
上下文无关语法是更宽松的。

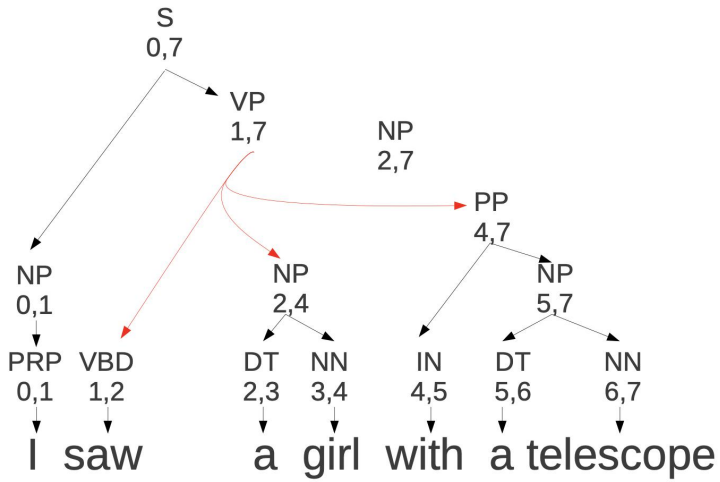
只看语法，不看常识中的语义。

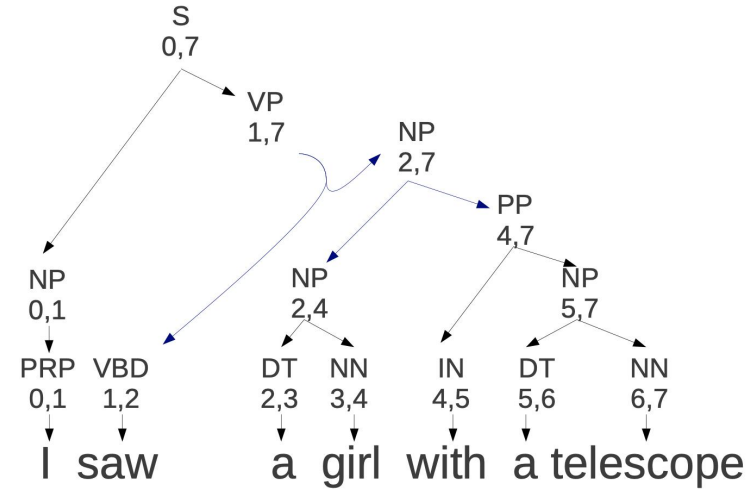
但是，往往，很多优秀的文学作品往往是“上下文无关”的...



歧义性




S
0,7
VP
1,7
NP
2,7
PP
4,7
NP
0,1
NP
2,4
NP
5,7
PRP
0,1
VBD
1,2
DT
2,3
NN
3,4
IN
4,5
DT
5,6
NN
6,7
I saw a girl with a telescope


S
0,7
VP
1,7
NP
2,7
PP
4,7
NP
0,1
NP
2,4
NP
5,7
PRP
0,1
VBD
1,2
DT
2,3
NN
3,4
IN
4,5
DT
5,6
NN
6,7
I saw a girl with a telescope

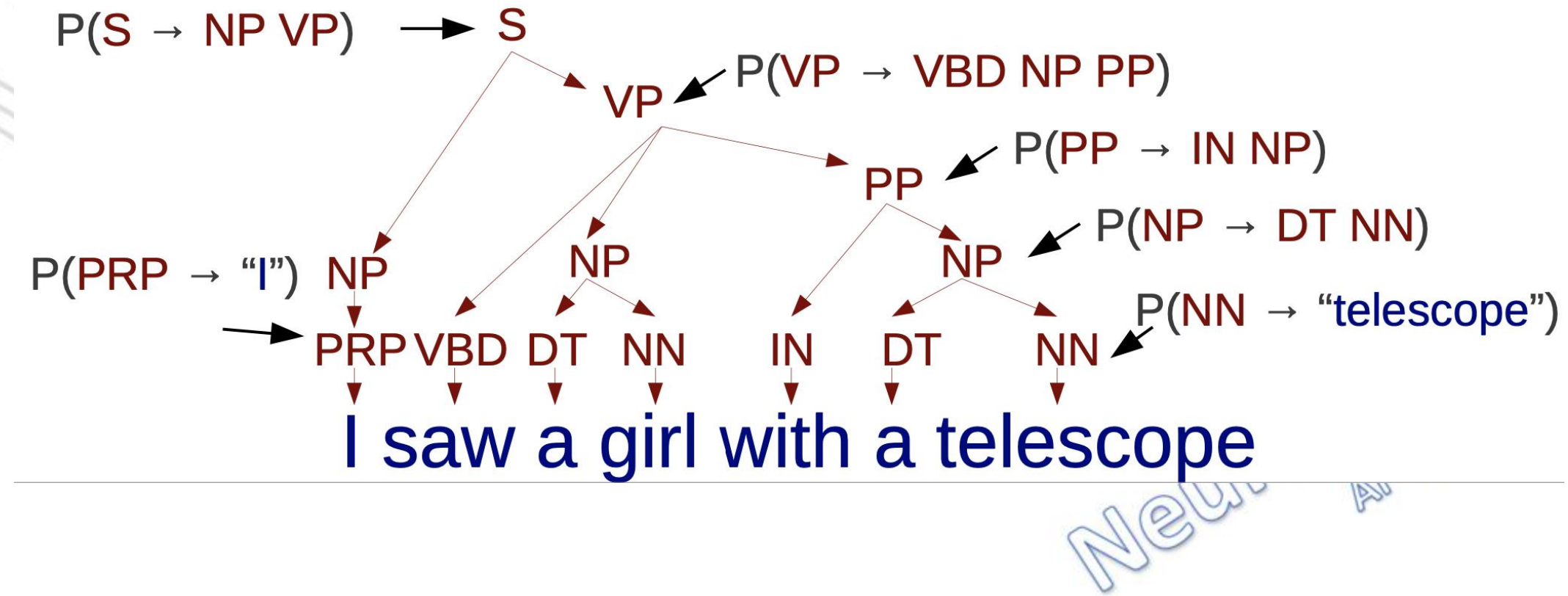
如何解决歧义性

Probabilistic Context Free Grammar (PCFG)

PCFG: Define probability for each node

NeuHub
AI Links All

如何解决歧义性


P(S → NP VP)
S
VP
P(VP → VBD NP PP)
PP
P(PP → IN NP)
NP
P(NP → DT NN)
NP
P(PRP → "I")
NP
P(NN → "telescope")
PRP VBD DT NN IN DT NN
I saw a girl with a telescope

如何解决歧义性

可以计算某种分析方式对应的生成概率

P(S → NP VP) * P(NP → PRP) * P(PRP → “I”)
* P(VP → VBD NP PP) * P(VBD → “saw”) * P(NP → DT NN)
* P(DT → “a”) * P(NN → “girl”) * P(PP → IN NP) * P(IN → “with”)
* P(NP → DT NN) * P(DT → “a”) * P(NN → “telescope”)

NeuHub AI Link

Chomsky Normal Form

给定文法规则及相应概率，如何分析某个句子的句法结构（得到句法结构树）？

NeuHub AI Link

Chomsky Normal Form

- All rules have two non-terminals or one terminal on right

OK	OK	Not OK!
S → NP VP	PRP → "I"	VP → VBD NP PP
S → PRP VP	VBD → "saw"	NP → NN
VP → VBD NP	DT → "a"	NP → PRP

- Can convert rules into CNF

VP → VBD NP PP ➡ VP → VBD VP'
VP' → NP PP

NP → PRP + PRP → "I" ➡ NP_PRP → "I"

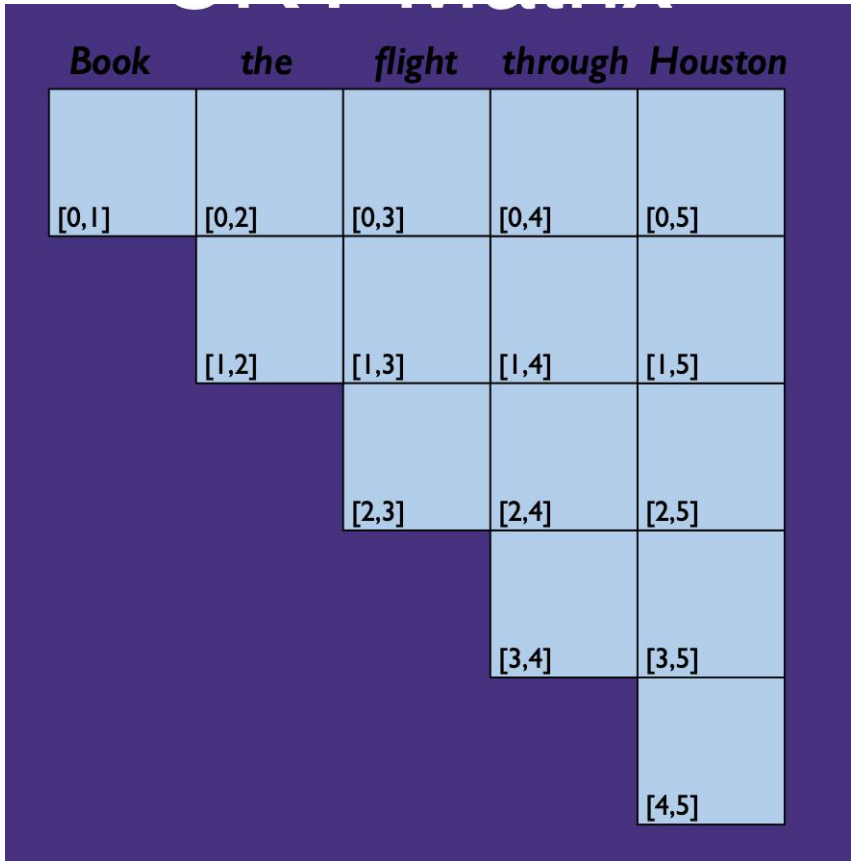
Chomsky Normal Form

$\mathscr{L}_1$ Grammar	$\mathscr{L}_1$ in CNF
$S \rightarrow NP\ VP$	$S \rightarrow NP\ VP$
$S \rightarrow Aux\ NP\ VP$	$S \rightarrow X1\ VP$
	$X1 \rightarrow Aux\ NP$
$S \rightarrow VP$	$S \rightarrow book \mid include \mid prefer$
	$S \rightarrow Verb\ NP$
	$S \rightarrow X2\ PP$
	$S \rightarrow Verb\ PP$
	$S \rightarrow VP\ PP$
$NP \rightarrow Pronoun$	$NP \rightarrow I \mid she \mid me$
$NP \rightarrow Proper\text{-}Noun$	$NP \rightarrow TWA \mid Houston$
$NP \rightarrow Det\ Nominal$	$NP \rightarrow Det\ Nominal$
$Nominal \rightarrow Noun$	$Nominal \rightarrow book \mid flight \mid meal \mid money$
$Nominal \rightarrow Nominal\ Noun$	$Nominal \rightarrow Nominal\ Noun$
$Nominal \rightarrow Nominal\ PP$	$Nominal \rightarrow Nominal\ PP$
$VP \rightarrow Verb$	$VP \rightarrow book \mid include \mid prefer$
$VP \rightarrow Verb\ NP$	$VP \rightarrow Verb\ NP$
$VP \rightarrow Verb\ NP\ PP$	$VP \rightarrow X2\ PP$
	$X2 \rightarrow Verb\ NP$
$VP \rightarrow Verb\ PP$	$VP \rightarrow Verb\ PP$
$VP \rightarrow VP\ PP$	$VP \rightarrow VP\ PP$
$PP \rightarrow Preposition\ NP$	$PP \rightarrow Preposition\ NP$

京东智联云 & 贪心科技 GREEDY TECHNOLOGY
AI LeuHub AI Link
贪心科技 | 让每个人享受个性化教育服务

CKY Parsing: A Dynamic Programming Approach

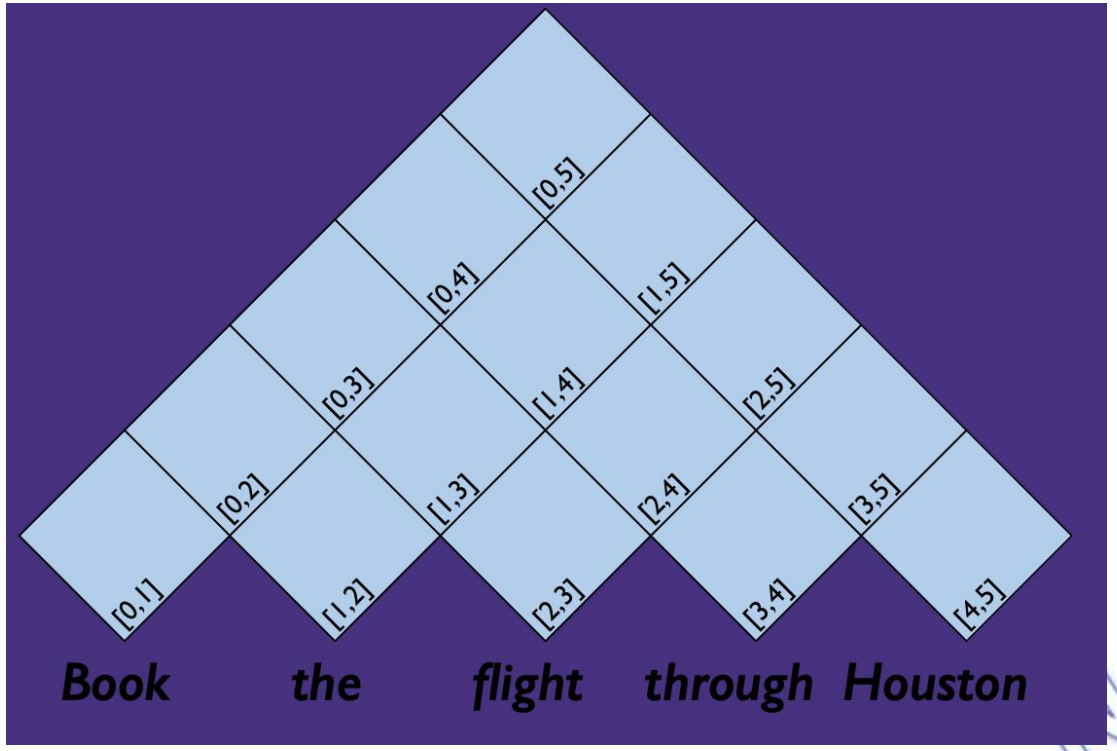
新建一个(n + 1) × (n + 1)
的矩阵（取右上三角）



NeuHub
AI Link

CKY Parsing: A Dynamic Programming Approach

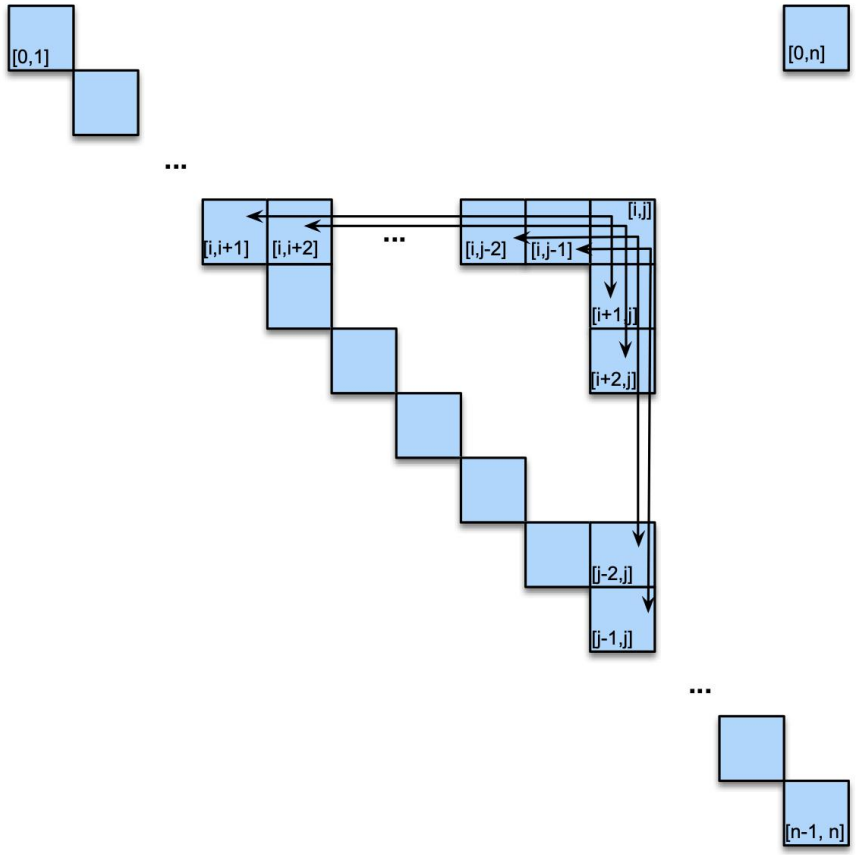
产生式：A → B C
对于节点A[i, j]，产生BC两个节点
B 必须在节点 [i, k]
C 必须在节点 [k, j]
i < k < j
j > i + 1



如何进行状态转移？

NeuHub AI Link

CKY Parsing: A Dynamic Programming Approach



如何进行状态转移？

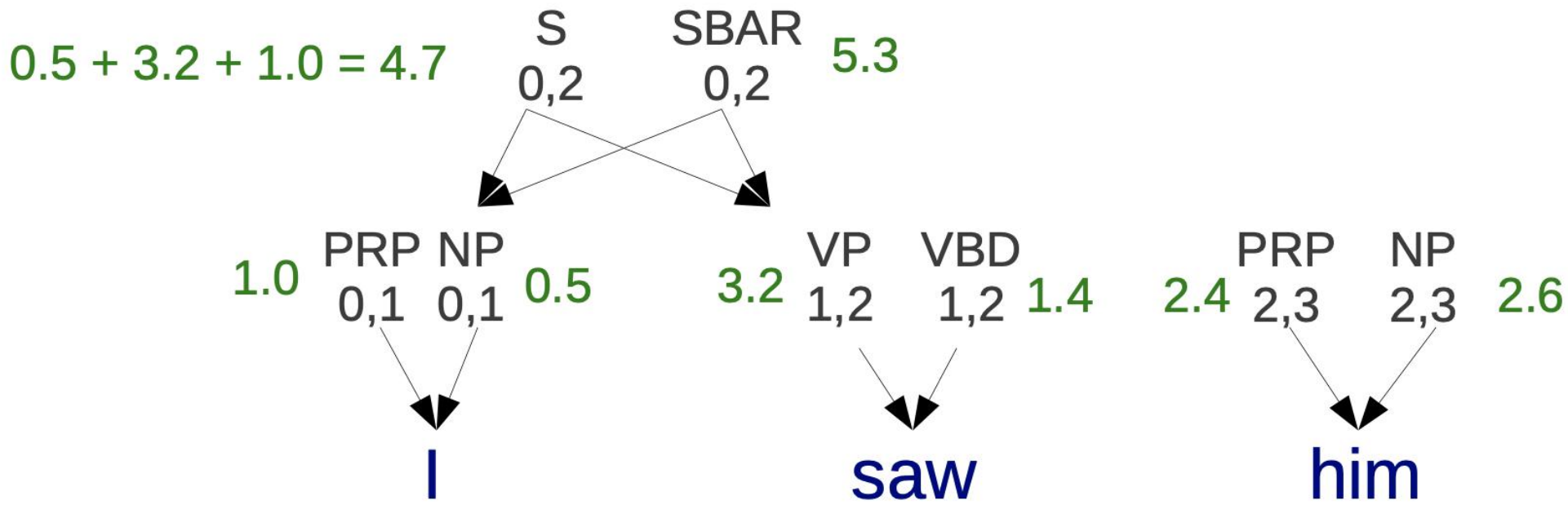
最后算出[0,n]的最大值并回溯出路径即得到语法树

贪心科技 | 让每个人享受个性化教育服务

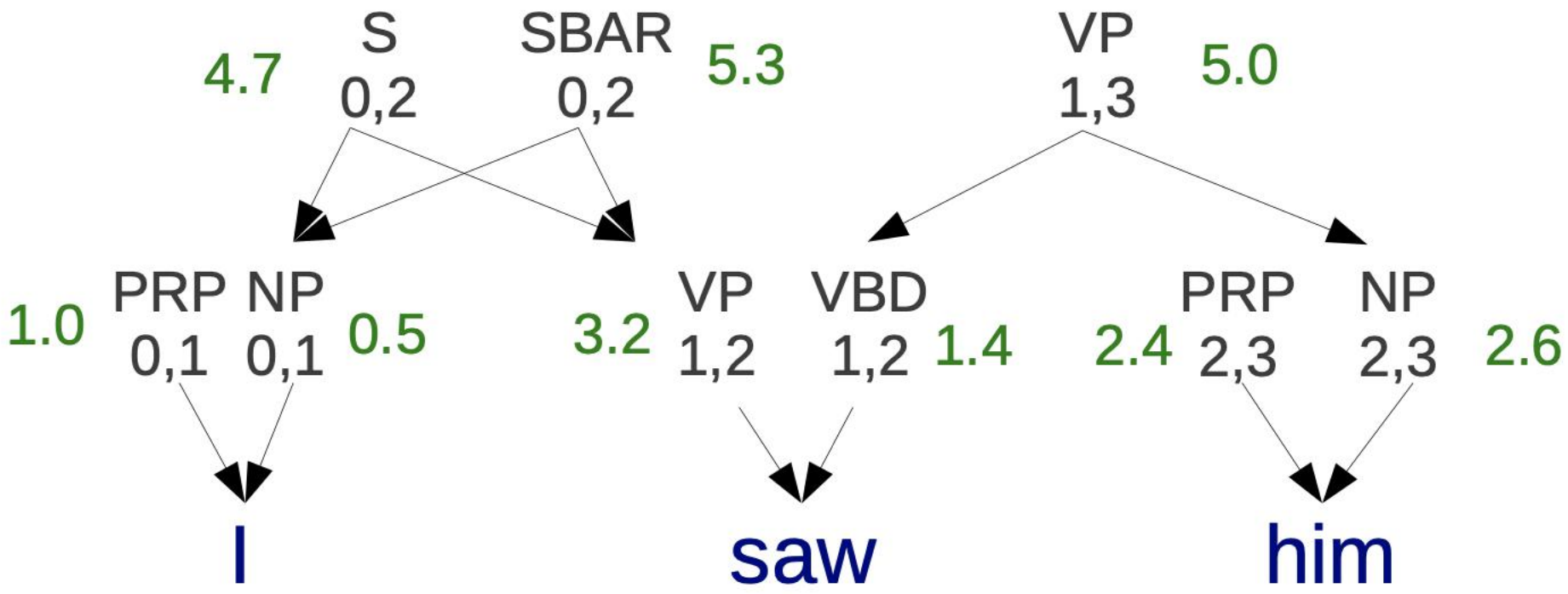
例子

这里的绿色值表示的是-log(P)，因此路径值越小越好
如：对于NP -> I 对应的-log(P)值为0.5

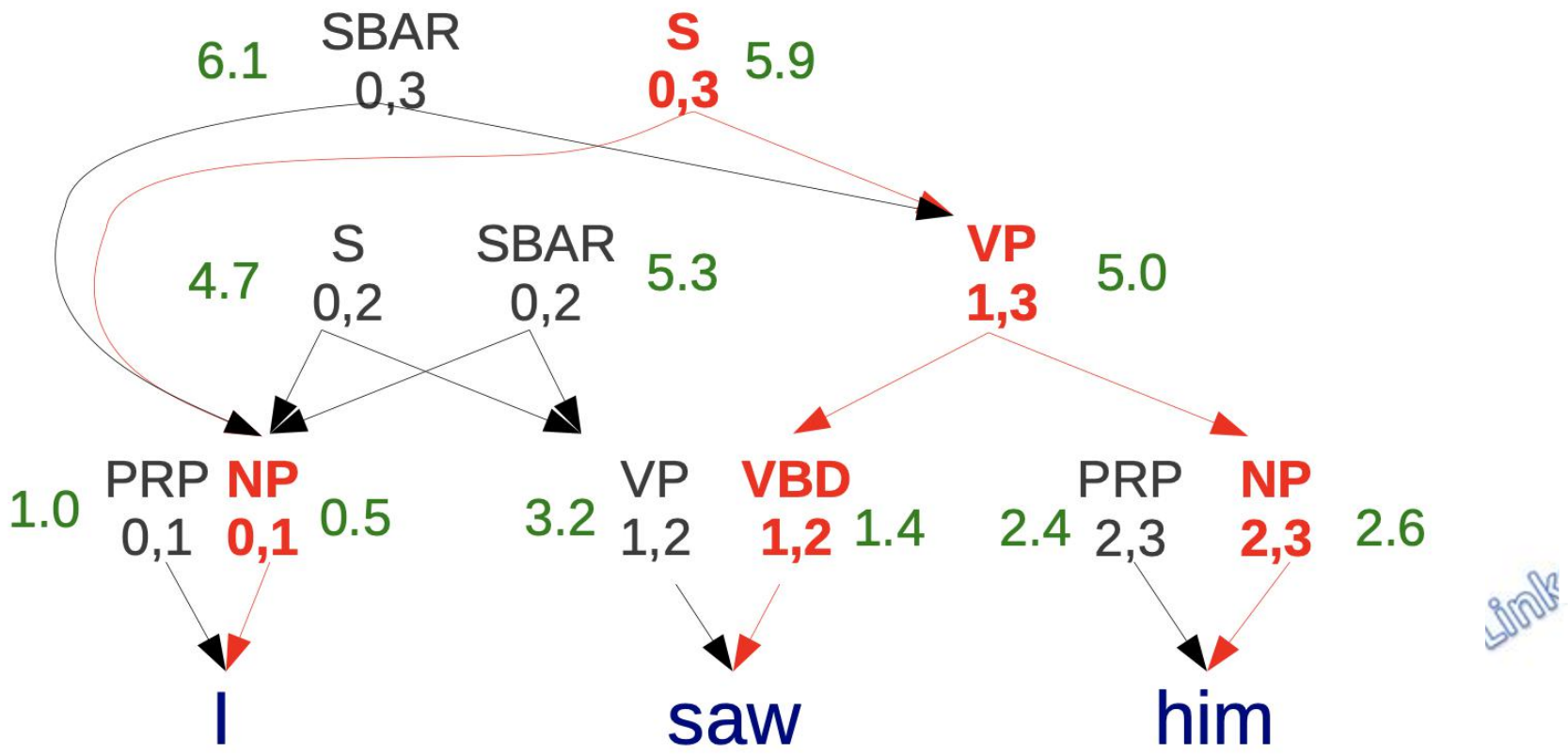
例子


0.5 + 3.2 + 1.0 = 4.7
S
0,2
SBAR
0,2
5.3
1.0
PRP
0,1
NP
0,1
0.5
3.2
VP
1,2
VBD
1,2
1.4
2.4
PRP
2,3
NP
2,3
2.6
I
saw
him

例子


4.7
S
0,2
SBAR
0,2
5.3
VP
1,3
5.0
1.0
PRP
0,1
NP
0,1
0.5
3.2
VP
1,2
VBD
1,2
1.4
2.4
PRP
2,3
NP
2,3
2.6
I
saw
him

例子


6.1
SBAR
0,3
S
0,3
5.9
4.7
S
0,2
SBAR
0,2
5.3
VP
1,3
5.0
1.0
PRP
0,1
NP
0,1
0.5
3.2
VP
1,2
VBD
1,2
1.4
2.4
PRP
2,3
NP
2,3
2.6
I
saw
him

回顾

- Context-free grammar的应用
- 如何理解上下文无关
- 歧义性以及如何解决歧义性
- Chomsky Normal Form是什么？
- 如何进行CKY Parsing（如何进行状态转移）

京东智联云 & 贪心科技 GREEDY TECHNOLOGY
NeuHub AI Links All

参考

http://www.phontron.com/slides/nlp-programming-en-10-parsing.pdf

https://www.shane.st/teaching/571/aut19/slides/3_CNF-CKY.pdf

http://qmznj.cn/index.php/2018/05/23/cyk-algorithm-introduction-and-implementation/

https://web.stanford.edu/~jurafsky/slp3/13.pdf

京东智联云 & 贪心科技

贪心科技 | 让每个人享受个性化教育服务

NeuHub AI Links All